Additive kappa can be increased by combining adjacent categories

Warrens, Matthijs J

Published in:
International Mathematical Forum

DOI:
10.12988/imf.2015.5431

International Mathematical Forum, Vol. 10, 2015, no. 7, 323 - 328
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/imf.2015.5431

Additive Kappa can be Increased by Combining Adjacent Categories

Matthijs J. Warrens

Institute of Psychology, Leiden University
Leiden, The Netherlands

Abstract

Weighted kappa is a measure that is commonly used for quantifying similarity between two ordinal variables with identical categories. Additive kappa is a special case of weighted kappa that allows the researcher to specify distances between adjacent categories. It is shown that additive kappa is a weighted average of the additive kappas of all collapsed tables of a specific size. It follows that, if the reliability of a categorical rating instrument is assessed with additive kappa, the reliability can be increased by combining categories.

Mathematics Subject Classification: 62H17, 62H20, 62P15, 62P25

Keywords: Ordinal similarity; Inter-observer reliability

1 Introduction

In pattern recognition, classification and data analysis, similarity measures are used to quantify the strength of a relationship between two variables. Commonly used examples are, Pearson's product-moment correlation for measuring linear dependence between two numerical variables, the Jaccard coefficient for measuring co-occurrence of two species types, and the Hubert-Arabie adjusted Rand index for comparing partitions of two different clustering algorithms. A commonly used coefficient for measuring similarity between two ordinal variables with identical categories is the weighted kappa measure [1-4]. Moreover, weighted kappa is a standard tool for assessing reliability of categorical rating instruments and scales in social, behavioral and medical sciences [2,4].

Suppose the data consist of two variables X and Y with identical ordinal categories $A_1, A_2, \ldots, A_c$. For example, if the rating scale would measure rigidity of a subject, the category labels could be, Absent, Slight, Moderate, Severe. The variables contain categorical scores of n subjects. For a population of subjects, let π_{ij} denote the proportion classified in category i on X and in category j on Y, where $i, j \in \{1, 2, \ldots, c\}$. Furthermore, define $\pi_{i+} := \sum_{j=1}^{c} \pi_{ij}$ and $\pi_{+i} := \sum_{j=1}^{c} \pi_{ji}$. The marginal probabilities π_{i+} and π_{+i} reflect how many subjects are in category A_i of X and Y, respectively.

With ordered categories dissimilarity between the variables on adjacent categories is usually of greater importance than dissimilarity on categories that are further apart. Weighted kappa allows the user to specify weights to describe the closeness between categories. If the weights are distances, pairs of categories that are further apart are assigned higher weights. Let $w_{ij} \geq 0$ for $i, j \in \{1, 2, \ldots, c\}$ be non-negative real numbers, with $w_{ii} = 0$. The weighted observed dissimilarity between the variables is defined as $O := \sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} \pi_{ij}$. Weighted kappa corrects for (dis)similarity due to chance. The expectation of the weighted dissimilarity under independence is given by $E := \sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} \pi_{i+} \pi_{+j}$. Cohen's weighted kappa is then defined as [1-4]

$$\kappa_w := 1 - \frac{O}{E} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} \pi_{ij}}{\sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} \pi_{i+} \pi_{+j}}. \tag{1}$$

Measure (1) is a function from the set of all square contingency tables to the interval $[1, -1]$. Its value is 1 when $O = 0$ $(X = Y)$, 0 when $O = E$, and negative when $O > E$. The fraction on the right-hand side of (1) shows that the value of (1) is invariant under multiplication of the weights w_{ij} by a positive constant.

Let $d_1, d_2, \ldots, d_{c-1} \geq 0$ be distances between the $c - 1$ pairs of adjacent categories. The additive weights are defined as

$$w_{ij} = 0 \;\text{ for } i = j; \quad w_{ij} = \sum_{k=i}^{j-1} d_k \;\text{ for } i < j; \quad w_{ij} = \sum_{k=j}^{i-1} d_k \;\text{ for } i > j. \tag{2}$$

The weight in (2) can be seen as a distance between categories A_i and A_j on an underlying one-dimensional interval scale. If category A_1 is the origin then the amounts d_k for $k \in \{1, 2, \ldots, c - 1\}$ indicate the relative locations of categories $A_2, A_3, \ldots, A_c$ respectively. The weights are called additive weights since additivity holds between these distances.

Substituting (2) into (1) we obtain additive kappa [3]

$$\kappa_a\left(d_1,\ldots,d_{c-1}\right)=1-\frac{\sum_{i=1}^{c-1}\sum_{j=i+1}^{c}\left(\sum_{k=i}^{j-1}d_k\right)\left(\pi_{ij}+\pi_{ji}\right)}{\sum_{i=1}^{c-1}\sum_{j=i+1}^{c}\left(\sum_{k=i}^{j-1}d_k\right)\left(\pi_{i+}\pi_{+j}+\pi_{j+}\pi_{+i}\right)}. \tag{3}$$

If we have $d_k=1$ for $k\in\{1,2,\ldots,c-1\}$ the weights in (2) are identical to the so-called linear weights [1,4]. A limitation of linear kappa is that the ordered categories are assumed to be equidistant, which is an unreasonable assumption for many ordinal variables in real life applications.

2 A weighted average

In reliability studies it is sometimes desirable to combine some of the categories and shorten the rating scale, for example, when two categories are easily confused [1]. With ordered categories it only makes sense to combine categories that are adjacent in the ordering. Theorem 2.1 below shows that the overall additive kappa is a weighted average of additive kappas of all collapsed tables of a specific size.

If the agreement table has c categories additive kappa requires the specification of the $c-1$ distances $d_1,d_2,\ldots,d_{c-1}$ for weighting scheme (2). If we combine categories the collapsed table has less categories and we have to specify a new set of distances between the categories. We will use the following rule. If we combine the categories A_k and A_{k+1} the distance d_k drops from the weighting scheme. The set of distances for the weighting scheme of the collapsed $(c-1)\times(c-1)$ table is then given by $d_1,\ldots,d_{k-1},d_{k+1},\ldots,d_{c-1}$. If we combine multiple categories at once all distances between the associated categories are dropped from the weighting scheme.

Suppose the agreement table has c categories and let $m\in\{2,\ldots,c-1\}$ be fixed. The agreement table of size $c\times c$ becomes an $m\times m$ table if we combine $c-m$ pairs of adjacent categories. Since we have $c-1$ pairs there are

$$M(c,m)=\binom{c-1}{c-m}=\binom{c-1}{m-1} \tag{4}$$

ways to choose $c-m$ from the $c-1$ pairs. Thus, the agreement table of size $c\times c$ can be collapsed into $M(c,m)$ distinct tables of size $m\times m$. With regard to Theorem 2.1 below let O_ℓ and E_ℓ for $\ell\in\{1,2,\ldots,M\}$ denote the observed and expected weighted disagreement of these $M(c,m)$ smaller tables. Furthermore, define

$$\kappa_\ell:=1-\frac{O_\ell}{E_\ell}\quad\text{for}\quad\ell\in\{1,2,\ldots,M\}. \tag{5}$$

The κ_ℓ denote the additive kappas of the M subtables.

Theorem 2.1 shows that the overall additive kappa κ_a is a weighted average of the additive kappas κ_ℓ of the subtables. The weights are the denominators E_ℓ of the weighted kappas. Theorem 2.1 generalizes the main results in [1,3]. The proof below consists of new arguments and provides more inside than the technical proof used in [1].

Theorem 2.1. *Consider an agreement table $\{\pi_{ij}\}$ with $c \geq 3$ categories and consider the M collapsed tables of size $m \times m$. We have*

$$\kappa_a = \frac{\sum_{\ell=1}^{M} E_\ell \kappa_\ell}{\sum_{\ell=1}^{M} E_\ell}. \tag{6}$$

Proof. Let O_a and E_a denote, respectively, the weighted observed dissimilarity and the expectation of the weighted dissimilarity under independence of weighted kappa κ_a. We first derive the identity

$$\sum_{\ell=1}^{M} O_\ell = N(c,m) \cdot O_a, \tag{7}$$

where

$$N(c,m) = \binom{c-2}{m-2}. \tag{8}$$

Consider an arbitrary element π_{ij} of $\{\pi_{ij}\}$. If $i = j$ we have $w_{ii} = 0$. Therefore, assume $i \neq j$. Since κ_a and the κ_ℓ are symmetric, the elements π_{ij} and π_{ji} have the same weights. Therefore, assume $i < j$. The weight of π_{ij} in O_a is the total distance between categories A_i and A_j, given in (2). If we combine two categories A_k and A_{k+1} the distance d_k drops from the weighting scheme and is not used in the calculation of the weights. For an $m \times m$ table the weight of π_{ij} is thus smaller than w_{ij}. If we consider all M tables of size $m \times m$ each distance d_k drops out the weighting scheme the same number of times. Hence, since we sum over all M tables in (7) it suffices to determine how often a specific distance d_k drops from the weighting scheme.

The number of times a distance d_k drops out the weighting scheme of an $m \times m$ table is given by

$$\binom{c-2}{c-m-1}, \tag{9}$$

which is the number of ways to choose $c-m-1$ distances from the remaining $c-2$ pairs of distances. Since

$$\binom{c-2}{c-m-1} + \binom{c-2}{m-2} = \binom{c-1}{m-1} = M(c,m), \tag{10}$$

the number of times a distance d_k is involved in the calculation of the weights of an $m \times m$ table is given in (8). Hence, if we sum over all M subtables of size $m \times m$ we obtain the identity in (7).

Next, applying similar arguments to the $c \times c$ table $\{\pi_{i+}\pi_{+j}\}$ and the E_ℓ, we obtain the identity

$$\sum_{\ell=1}^{M} E_\ell = N(c,m) \cdot E_a. \tag{11}$$

Finally, using (7) and (11), together with the identity

$$E_\ell \kappa_\ell = E_\ell \left(1 - \frac{O_\ell}{E_\ell}\right) = E_\ell - O_\ell, \tag{12}$$

we have

$$\frac{\sum_{\ell=1}^{M} E_\ell \kappa_\ell}{\sum_{\ell=1}^{M} E_\ell} = \frac{\sum_{\ell=1}^{M} (E_\ell - O_\ell)}{\sum_{\ell=1}^{M} E_\ell} = \frac{N(c,m) \cdot E_a - N(c,m) \cdot O_a}{N(c,m) \cdot E_a} = \kappa_a. \tag{13}$$

□

3 Conclusion

Theorem 2.1 shows that the overall additive kappa is a weighted average of additive kappas of all collapsed tables of a specific size. Theorem 2.1 shows in particular that the additive kappa of an $c \times c$ table is a weighted average of the additive kappas of all $(c-1) \times (c-1)$ tables that are obtained by combining two adjacent categories. If the data do not have a particular structure [5] then these additive kappas are all distinct. This implies that there in general exist two categories such that, when combined, additive kappa increases. In addition, there exist two categories such that, when combined, additive kappa decreases. Theorem 2.1 thus implies an existence result. Moreover, if we measure inter-observer reliability in terms of additive kappa, the reliability can thus be increased by shortening the rating scale.

References

[1] M. J. Warrens, Cohen's linearly weighted kappa is a weighted average, *Advances in Data Analysis and Classification,* **6** (2012), 67 - 79. http://dx.doi.org/10.1007/s11634-011-0094-7

[2] M. J. Warrens, Some paradoxical results for the quadratically weighted kappa, *Psychometrika,* **77** (2012), 315 - 323. http://dx.doi.org/10.1007/s11336-012-9258-4

[3] M. J. Warrens, Cohen's weighted kappa with additive weights, *Advances in Data Analysis and Classification,* **7** (2013), 41 - 55. http://dx.doi.org/10.1007/s11634-013-0123-9

[4] M. J. Warrens, Conditional inequalities between Cohen's kappa and weighted kappas, *Statistical Methodology,* **10** (2013), 14 - 22. http://dx.doi.org/10.1016/j.stamet.2012.05.004

[5] M. J. Warrens, On agreement tables with constant kappa values, *Advances in Statistics,* (2014), ID 853090. http://dx.doi.org/10.1155/2014/853090

Received: April 24, 2015; Published: May 15, 2015